

Distribuciones y simulación. El caso de la verificación de un dado

Vicente D. Estruch^a, Francisco J. Boigues^a y Anna Vidal^a

^aInstituto de Investigación para la Gestión Integrada de Zonas Costeras, Departamento de Matemática Aplicada, Universitat politècnica de València, (vdestruc@mat.upv.es, fraboipl@mat.upv.es y avidal@mat.upv.es).

Abstract

From the perspective of mathematical statistics, a random variable is a purely theoretical concept, but it can be focused from an approach based on the construction of a random mathematical model. The random variable, as its name suggests, takes its values in a completely random as if the results of a "lottery" were involved. However the random variable is subject to a law of chance, which becomes noticeable if the "lottery" associated with the behavior of the random variable is repeated a lot of times. This paper describes the development of a learning experience of the concept of random variable by using simulation and experimentation in the classroom, with the assistance of the flip teaching methodology.

Keywords: Flip teaching, discrete random variable, modeling, simulation, Chi-square test, Microsoft Excel[®], Matlab[®]

Resumen

Desde la perspectiva de la estadística matemática, el concepto de variable aleatoria es puramente teórico, pero se puede enfocar desde un planteamiento basado en la construcción de un modelo matemático aleatorio. La variable aleatoria, como su nombre indica, toma sus valores de forma totalmente aleatoria, como si de los resultados de una "lotería" se tratara. Sin embargo, la variable aleatoria está sujeta a una ley de probabilidad, que se evidencia si se repite muchas veces la "lotería" asociada al comportamiento de la variable aleatoria. Este trabajo describe el desarrollo de una experiencia de aprendizaje del concepto de variable aleatoria mediante simulación y experimentación en el aula, con la ayuda de la metodología "flip teaching".

Palabras clave: Flip teaching, variable aleatoria discreta, modelización, simulación, test Chi-cuadrado, Microsoft Excel[®], Matlab[®].

1. Introducción

1.1. La estadística en el contexto universitario

En la formación universitaria que facilita a los alumnos la adquisición de competencias para el ejercicio de profesiones relacionadas con las ciencias de la naturaleza o ingenierías, se incluye la estadística como parte importante de las materias básicas. En dicha formación se consideran aspectos teóricos de la estadística, aunque se suele destacar sobre todo su carácter instrumental, orientado a la resolución práctica de problemas en los que hay que

manejar conjuntos relativamente grandes de datos, por lo que el uso de paquetes estadísticos y, por lo tanto del ordenador, tendrá un papel esencial.

Por otra parte, a la luz de investigaciones recientes, el aprendizaje efectivo de la estadística en la universidad necesita sustentarse en la comprensión de conceptos estadísticos básicos y en la construcción progresiva de representaciones mentales implícitas o explícitas. (Escalante Gómez, 2008), lo cual exige abordar en el aula, con rigor, las bases teóricas que sustentan las técnicas de análisis estadístico. En este punto aparecen problemas de diversa índole, donde destaca la muy escasa formación estadística de los alumnos de bachillerato o de ciclos formativos que acceden a la universidad, no siendo menos importante el problema actitudinal. Los déficits motivacionales y actitudinales constatados, así como la importancia concedida a la dimensión actitudinal en cualquier desempeño competente, han alentado el desarrollo de un área de investigación bien definida, y relativamente reciente, que aborda el dominio afectivo-actitudinal en los procesos de enseñanza-aprendizaje de la Estadística (Blanco, 2008).

1.2. Los métodos probabilísticos desde la perspectiva de los problemas reales

Las matemáticas que necesitaremos para afrontar los problemas relacionados con resultados de experimentos aleatorios constituyen la referencia que lleva a clasificar los métodos probabilísticos en discretos y continuos. Un enfoque discreto se utiliza cuando el número de resultados experimentales es finito o infinito numerable. Por otro lado, en un experimento en que se mide, por ejemplo, el tiempo que tarda un barco oceanográfico en tener un avistamiento de cierta especie de ballena, es posible que los resultados puedan variar desde 0 a T horas. En este caso los resultados posibles constituyen un conjunto infinito no numerable, el de los valores reales que forman el intervalo $[0, T]$, lo que nos lleva a un enfoque continuo.

En el enfoque discreto, es posible asignar probabilidades a los resultados individuales. En el enfoque continuo no será posible asignar una probabilidad no nula a cada resultado, y sólo podremos asignar probabilidades a sucesos asociados a que la variable pertenezca a un subintervalo del intervalo total de los valores posibles. Siguiendo con el ejemplo del avistamiento de ballenas, podríamos plantear el cálculo de la probabilidad de que se produzca un avistamiento al pasar un tiempo situado dentro de los límites del intervalo temporal $[a,b] \subset [0,T]$.

El caso discreto finito es el más fácil de conceptualizar y de describir matemáticamente, lo cual facilita obtener ejemplos ilustrativos que permitan abordar con éxito una primera etapa en el estudio formal de las variables aleatorias.

1.3. La probabilidad en el caso discreto

Supongamos un experimento aleatorio en el que se pueden obtener los posibles resultados numéricos $\{X_1, X_2, \dots, X_n\}$. Entonces la suma de las probabilidades de cada suceso, $P(X_i) \geq 0$, $i=1,2,\dots,n$, ha de ser 1, es decir $P(X_1)+P(X_2)+\dots+P(X_n)=1$. Si el número de posibles resultados $\{X_1, X_2, \dots, X_i, \dots\}$ es infinito numerable, entonces $\sum_{i=1}^{+\infty} P(X_i) = 1$, lo que conlleva que necesitemos recurrir al concepto de serie numérica y la función de

probabilidad $P(X)$ solo tendrá sentido si la serie de números no negativos $\sum_{i=1}^{+\infty} P(X_i)$ es convergente.

En el caso particular de que todos los resultados tengan la misma probabilidad de suceder, $P(X_i) = k > 0$, el número de resultados ha de ser finito ya que la serie $\sum_{i=1}^{+\infty} k$, $k > 0$, siempre es divergente. En el caso de un número finito de resultados del experimento aleatorio $\{X_1, X_2, \dots, X_n\}$, todos con la misma probabilidad, según la definición de probabilidad de Laplace, la probabilidad de que suceda cualquiera de ellos será $P(X_i) = \frac{1}{n}$, $i=1, 2, \dots, n$.

1.4. El problema y el modelo

Los modelos de probabilidad son aproximaciones simplificadas de la realidad. Como tales aproximaciones, deben contemplar las características más importantes del fenómeno aleatorio real para ser útiles como herramientas de predicción, teniendo en cuenta que, por otra parte, deben ser lo más simples posible, ya que un modelo complejo difícil de manejar no será útil para ser utilizado en la práctica.

Consideremos la siguiente situación real (R1):

R1: *Datos empíricos obtenidos mediante observación, indican que cierto roedor se reproduce de forma que, en el momento del alumbramiento, en cada camada nos encontramos con 1, 2, 3, 4, o 6 crías vivas con la misma proporción, siendo las proporciones de camadas nulas (con ninguna cría viva) y de camadas con un número de crías vivas superior a 6 prácticamente despreciables.*

Es obvio que si hacemos la simplificación de considerar nula la probabilidad de 0 crías vivas y la de más de 6 crías vivas en la camada, la situación planteada es asimilable al experimento aleatorio de lanzar un dado perfecto y observar el valor que aparece en la cara superior (R2):

R2: *En el experimento aleatorio de lanzar un dado, en la cara superior puede aparecer cualquiera de los valores $\{1, 2, 3, 4, 5, 6\}$, con la misma probabilidad.*

Ambas situaciones, la del número de alumbramientos en la camada y el lanzamiento del dado, responden al mismo modelo teórico (M) que se describe a continuación:

M: *Sea X una variable aleatoria que toma los valores $\{1, 2, 3, 4, 5, 6\}$, con la misma probabilidad, dada por la función $f(i) = P(X=i) = \frac{1}{6}$, $i=1, 2, 3, 4, 5, 6$. En este caso se dice que la variable aleatoria X sigue una distribución uniforme discreta, o rectangular discreta, en el intervalo $[1, 6]$.*

Es fácil calcular el valor esperado (media) y la varianza de la variable X definida en **M**: $E(X) = \mu = 3.5$ y $\sigma^2 = 2.91\bar{6}$, respectivamente.

1.5. Análisis teórico frente a la simulación por ordenador

La situación que nos ocupa es fácil de abordar desde el punto de vista teórico ya que conocer con detalle la variable aleatoria discreta X , que puede tomar los valores $\{1, 2, 3, 4,$

5, 6}, cuya función de probabilidad viene dada por $f(x)=1/6$, $x \in \{1,2,3,4,5,6\}$, permite calcular de forma inmediata probabilidades de interés u otras características estadísticas como el valor esperado o la desviación típica, por ejemplo. Sin embargo, a diferencia de los casos $R1$ y $R2$, en muchos problemas reales no es fácil determinar la función de probabilidad asociada a un modelo (o variable aleatoria) que representa el caso real. No obstante, si conocemos las características del fenómeno aleatorio, entonces podríamos llevar a cabo el experimento, o simularlo, un gran número de veces para obtener probabilidades aproximadas.

Por ejemplo, para determinar la probabilidad de obtener 3 caras al lanzar 4 veces una moneda perfecta, es decir con probabilidad de cara igual a 0.5, podemos lanzar la moneda cuatro veces y contar el número de caras (C). Supongamos que obtenemos $C_1=2$ caras. A continuación, repetimos el experimento lanzando la moneda cuatro veces más, obtenemos $C_2=1$ cara. De esta manera repetimos el experimento, por ejemplo 1000 veces, obteniéndose como resultado un conjunto finito (una muestra) de 1000 valores correspondientes al número de caras $\{C_1, C_2, \dots, C_{1000}\}$. Finalmente, para determinar la probabilidad de obtener 3 caras recurrimos a una aproximación de la definición de la probabilidad como límite de frecuencias relativas, lo cual nos proporciona el resultado (1)

$$P(3 \text{ caras al lanzar } 4 \text{ veces}) \approx \frac{n^{\circ} \text{ de veces en que se han obtenido } 3 \text{ caras}}{1000} \quad (1)$$

En los inicios de los estudios de probabilidad, los investigadores hacían exactamente este tipo de cálculos, obviamente muy tediosos por el hecho de tener que repetir muchas veces el experimento aleatorio. Resulta, pues, de suma importancia ser capaces de simular los procedimientos, es decir, siguiendo con el ejemplo de la moneda, obtener los resultados sin tener que realizar el trabajo físico de lanzar la moneda. Los ordenadores hacen que esto sea posible. Un ordenador permite realizar un cálculo miles o millones de veces. Lo único que se necesita es un medio (un programa, un lenguaje de programación y la correspondiente rutina) para simular el proceso. Afortunadamente la mayoría de los paquetes de software científicos incorporan generadores de números pseudo-aleatorios. Por ejemplo, Matlab, puede generar un valor, x , en el intervalo $(0,1)$ mediante la sentencia $x=\text{rand}(1,1)$. El número x es elegido, teóricamente, al azar por lo que pertenecerá a cualquier parte del intervalo $(0,1)$. Para simular el lanzamiento de una moneda, puesto que la probabilidad de que el valor x generado mediante $x=\text{rand}(1,1)$ pertenezca al intervalo $(0,1/2]$ será 0.5, entonces si $x \in (0,1/2]$, contabilizaremos cara y si $x \in (1/2,1)$, lo cual sucede también con una probabilidad 0.5, contabilizaremos cruz.

El lanzamiento de un dado, y por lo tanto también la aproximación al problema de las crías en una camada, puede hacerse mediante un ordenador. Para este caso, por ejemplo, tras generar el valor $x \in (0,1)$, se procedería del siguiente modo:

Algoritmo 1

Si $x \in (0,1/6]$ contabilizamos 1, si $x \in (1/6,2/6]$ contabilizamos 2, si $x \in (2/6,3/6]$ contabilizamos 3, si $x \in (3/6,4/6]$ contabilizamos 4, si $x \in (4/6,5/6]$ contabilizamos 5 y, por último, si $x \in (5/6,1)$ contabilizamos 6.

La simulación por ordenador de fenómenos aleatorios es una herramienta indispensable en la investigación científica moderna. Los métodos de Monte Carlo, caracterizados por basarse en generar valores aleatorios para simular procesos reales, también pueden utilizarse para facilitar a los estudiantes una mejor comprensión de los problemas probabilísticos.

1.6. Valores aleatorios y pseudo aleatorios

Es importante distinguir entre generar valores aleatorios y valores pseudo-aleatorios que, supuestamente, siguen cierta distribución.

Se define generador de números aleatorios como: “[...] un dispositivo informático o físico diseñado para producir secuencias de números sin un orden aparente.” (COLABORADORES DE WIKIPEDIA, Generador de números aleatorios). Por otra parte: “Un generador pseudoaleatorio de números (GPAN) es un algoritmo que produce una sucesión de números que es una muy buena aproximación a un conjunto aleatorio de números. La sucesión no es exactamente aleatoria en el sentido de que queda completamente determinada por un conjunto relativamente pequeño de valores iniciales,[...]. Si bien es posible generar sucesiones mediante generadores de números aleatorios por dispositivos mecánicos que son mejores aproximaciones a una sucesión aleatoria, los números pseudo-aleatorios son importantes en la práctica para simulaciones (por ejemplo, de sistemas físicos mediante el método de Montecarlo).” (COLABORADORES DE WIKIPEDIA, "Generador de números pseudoaleatorios")

Es importante remarcar que los paquetes de cálculo proporcionan realmente series de valores pseudo-aleatorios con lo que “se requiere de un cuidadoso análisis matemático para tener algún tipo de confianza en que un dado GPAN genera números que son suficientemente ‘aleatorios’ para ser útiles para el propósito para el que se los precisa.” (COLABORADORES DE WIKIPEDIA, "Generador de números pseudoaleatorios")

Una forma de verificar si los resultados de la simulación, donde se han realizado N imitaciones del experimento aleatorio, se ajustan al modelo teórico lo proporciona el estadístico Chi-cuadrado, $\chi^2 = \sum_i \frac{(N_i - E_i)^2}{E_i}$, donde N_i es la frecuencia absoluta del i -ésimo valor producto de la simulación y E_i es la frecuencia esperada para el modelo teórico de dicho valor. Se tiene que $E_i = N \cdot P_i$, donde P_i es la probabilidad del i -ésimo valor. En el caso de la variable aleatoria uniforme discreta [1,6], se tiene que $P_i = 1/6$, $i=1,2,3,4,5,6$. Para verificar si las simulaciones se ajustan al modelo teórico, elaboraremos la tabla 1.

Tabla 1. Modelo de tabla para el análisis de resultados

Resultado	Frecuencia absoluta en la simulación (N_i)	Probabilidad Teórica (P_i)	Frecuencia absoluta esperada ($N \cdot P_i$)	$C_i = \frac{(N_i - N/6)^2}{N/6}$
1	N_1	1/6	$N/6$	C_1
2	N_2	1/6	$N/6$	C_2
3	N_3	1/6	$N/6$	C_3
4	N_4	1/6	$N/6$	C_4
5	N_5	1/6	$N/6$	C_5
6	N_6	1/6	$N/6$	C_6

Y calculamos el valor de χ^2 (2)

$$\chi^2 = \sum_{i=1}^6 C_i = \sum_{i=1}^6 \frac{(N_i - N \cdot P_i)^2}{N_i \cdot P_i} = \sum_{i=1}^6 \frac{(N_i - \frac{N}{6})^2}{\frac{N}{6} \cdot \frac{6}{N}} = \sum_{i=1}^6 \left(N_i - \frac{N}{6}\right)^2 \quad (2)$$

Hay que tener en cuenta que los resultados de una simulación no son más que una muestra aleatoria y, por lo tanto, las conclusiones a las que lleguemos se basarán en la información parcial que proporciona dicha muestra, por lo que toda inferencia tendrá un nivel de confianza o vendrá avalada por un p -valor. A la hora de comparar entre generadores, se ajustará mejor al modelo teórico M , en base a las muestras obtenidas, aquel generador con un valor de χ^2 menor. Observemos que la expresión χ^2 no es más que una cuantificación de en qué medida se desvían las frecuencia absolutas obtenidas de la simulación respecto de las frecuencias teóricas esperadas. El valor χ^2 puede ser utilizado para contrastar la hipótesis de si la simulación se ajusta al modelo teórico o no (con un nivel de confianza prefijado) El estadístico muestral χ^2 , en el caso que nos ocupa, se distribuye Chi-cuadrado con $n-1$ grados de libertad, siendo n el número de categorías que comparamos. Al comparar el valor obtenido χ^2 con el valor teórico $\chi_{6-1,0.05}^2 = \chi_{5,0.05}^2 = 11.070$, correspondiente una variable aleatoria Y distribuida Chi-cuadrado con 4 grados de libertad, de forma que $P(Y \geq \chi_{5,0.05}^2) = 0.05$; entonces si $\chi^2 \leq 11.070$ podemos aceptar, a un nivel de confianza del 95%, que la simulación ha proporcionado una muestra que se ajusta al modelo teórico. En caso contrario ($\chi^2 > 11.070$) no podemos aceptar, a un nivel de confianza del 95%, que la simulación haya proporcionado una muestra que se ajuste al modelo teórico.

1.7. Un recorrido de estudio e investigación (REI) para aprender el concepto de variable aleatoria discreta y su ajuste a un modelo teórico: La verificación de la validez de un generador de valores (pseudo) aleatorios.

Se plantea un recorrido de estudio e investigación (REI) para la introducción del concepto de variable aleatoria discreta en la asignatura Instrumentos de Estadística y Simulación, de segundo curso de los estudios del Grado en Ciencias Ambientales en la Escuela Politécnica Superior de Gandia-Universitat Politècnica de València.

El REI que se propone seguiría los pasos que se enumeran a continuación:

1. Partimos de la situación real $R1$.
2. Se constata que, con las simplificaciones asumidas, la situación real $R1$ es análoga a la situación real $R2$ (se evidencia la necesidad de simplificar como paso previo a modelar).
3. Teniendo en cuenta la definición de variable aleatoria discreta, se constata que las situaciones reales $R1$ y $R2$ son dos concreciones del modelo teórico M .
4. Se utiliza un dado físico como generador de 1000 valores aleatorios y se analiza el nivel de ajuste al modelo teórico mediante el uso de un estadístico adecuado.
5. Se utiliza el programa EXCEL para generar 1000 valores pseudo-aleatorios para una distribución uniforme discreta en el intervalo $[1,6]$ y se analiza el nivel de ajuste al modelo teórico mediante el uso de un estadístico adecuado.

6. Se utiliza el programa MATLAB para generar 1000 valores pseudo-aleatorios para una distribución uniforme discreta en el intervalo [1,6] y se analiza el nivel de ajuste al modelo teórico mediante el uso de un estadístico adecuado.
7. Se valora qué generador de valores (pseudo) aleatorios se ajusta más al modelo teórico y si los generadores de valores (pseudo) aleatorios pueden ser aceptados como simuladores válidos de dicho modelo.

Los pasos del proceso se resumen en la figura 1

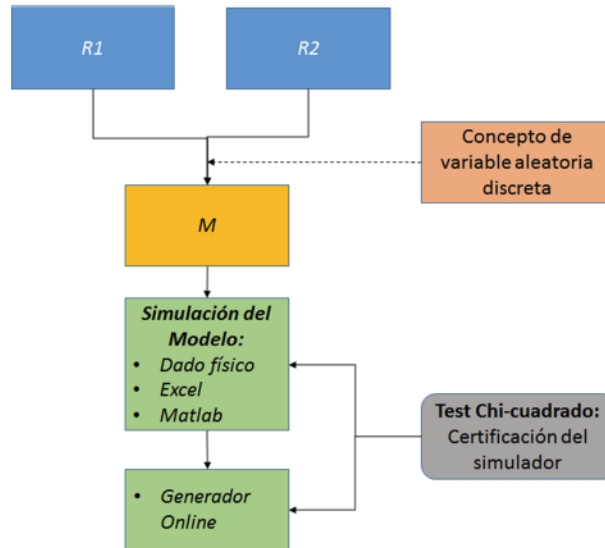


Fig.1. Itinerario del recorrido de estudio e investigación

2. Objetivos

El objetivo del desarrollo práctico del REI descrito en el punto 1.7 es que los alumnos adquieran las siguientes competencias:

- Distinguir entre modelo probabilístico y resultados de un experimento aleatorio.
- Modelar un fenómeno aleatorio sencillo.
- Utilizar diversos programas informáticos y aplicaciones online para obtener series de valores pseudo aleatorios.
- Analizar y establecer qué generadores de valores (pseudo) aleatorios proporcionan resultados (simulaciones) más ajustados a un modelo teórico y verificar si dichos generadores pueden ser aceptados como simuladores del modelo teórico.

3. Desarrollo de la innovación

La actividad se desarrollará inicialmente utilizando la técnica de la clase inversa (*flip teaching*). Comienza la actividad indicando a los alumnos que estudien, antes de la clase presencial, los apuntes proporcionados, que contesten a una serie de cuestiones planteadas y que visualicen una serie de videos sugeridos. Posteriormente en aula informática, en sesión única de 2 horas, los alumnos se organizan en grupos de dos por cada ordenador y realizan

una serie de experiencias de simulación. Finalmente se propone a cada grupo una tarea no presencial. El trabajo se organiza por etapas tal y como se describe a continuación.

Etapla 1.-No presencial

- a) Se expone el caso real *R1*. Planteando las cuestiones adecuadas, los alumnos llegan a la conclusión de que el caso *R1* se comporta de manera análoga al experimento aleatorio de lanzar un dado sobre una superficie plana y registrar el valor que aparece en la cara superior (*R2*).
- b) Teniendo en cuenta la definición de variable aleatoria discreta, que los alumnos habrán estudiado previamente, con el apoyo de los vídeos: *FisicayMates. Variables aleatorias discretas y continuas | Estadística UNED, FisicayMates. Variable Aleatoria Discreta | Función de probabilidad y de distribución*, llegan a comprender la formalización de los casos *R1* y *R2* mediante una variable aleatoria que se distribuye uniforme discreta en el intervalo [1,6].
- c) Se propone la visualización del vídeo *Estadística útil. ¿Qué es una prueba de bondad de ajuste?*, y la página web INSTITUTO TECNOLÓGICO DE CHIHUAHUA. *Prueba chi-cuadrada para la bondad del ajuste*, los cuales facilitan el aprendizaje de la prueba de bondad de ajuste basada en la distribución Chi-cuadrado: Las hipótesis que se manejan, el estadístico de prueba y la regla de decisión.
- d) Los alumnos deberán contestar a un cuestionario-examen PoliformaT, activo hasta el día anterior a la clase presencial.

Las etapas 2, 3 4 y 5 se desarrollan de forma presencial, en aula informática, en sesión de 2 horas.

Etapla 2.-.Simulación mediante un dado físico

Al principio de la clase presencial, durante 15 minutos, se comentan los resultados del cuestionario-examen y se atienden posibles dudas.

Un dado físico sin defectos y lanzado buscando un resultado al azar, es un excelente generador de valores aleatorios con distribución uniforme discreta en el intervalo [1,6].

Se proporcionará a cada grupo de dos personas un dado (figura 2 a)) y un fichero Excel con formato para registrar los resultados. Se pedirá a cada grupo de dos alumnos que lance 100 veces el dado y que registre los resultados. Se indicará a los alumnos cómo definir fórmulas y realizar gráficos en Excel de forma que, tras registrar los resultados de los 100 lanzamientos, se obtenga una tabla resumen con las frecuencias absolutas y relativas, y un gráfico que permita comparar las frecuencias relativas obtenidas en el lanzamiento del dado con las probabilidades teóricas del modelo, que se representa en el gráfico mediante una línea horizontal. También se les indicará cómo calcular el valor χ^2 a partir de los datos de la tabla (figura 2 b)).

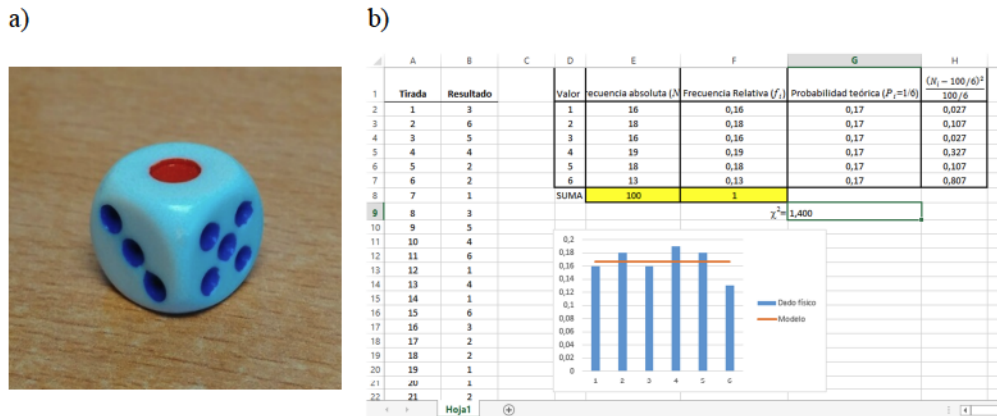


Fig. 2. a) Tipo de dado utilizado en la simulación. b) Hoja de registro de resultados con tabla de frecuencias absolutas y relativas, y grafico descriptivo.

En el ejemplo descrito en la figura 2 b), correspondiente a una simulación realizada por un grupo en clase, se ha obtenido que $\chi^2=1.4 < \chi^2_{5,0.05}=11.070$. Al comparar χ^2 con $\chi^2_{5,0.05}$, no sólo se cumple la desigualdad requerida sino que el valor de χ^2 es mucho menor que $\chi^2_{5,0.05}$, con lo que podemos afirmar, con un nivel de confianza del 95%, que la simulación nos ha proporcionado una muestra que se ajusta al modelo teórico; o lo que lo mismo, que la simulación con el dado físico ofrece resultados acordes con lo esperable para un dado perfecto.

Se pide a cada grupo que haga sólo 100 lanzamientos porque dicho trabajo es tedioso. Para convertir el trabajo en una experiencia colaborativa, se obtendrá una muestra de 1000 resultados agregando los resultados de 10 grupos de 2 personas. Los resultados de los diferentes grupos incorporan a una tabla global, que aparece proyectada en una pantalla, para que los resultados de los distintos grupos estén a disposición de todos. Los resultados obtenidos en una experiencia real en clase se exponen en la Tabla 2.

Tabla 2 Resultados de 10 grupos. En las dos últimas columnas se indican las frecuencias absolutas y relativas (Fa. y Fr.), respectivamente.

Resultado	Grupo										Fa.	Fr.
	1	2	3	4	5	6	7	8	9	10		
1	14	15	22	14	23	24	14	14	17	11	169	0,17
2	14	11	16	19	22	15	17	12	17	16	159	0,16
3	22	14	15	17	13	13	13	22	9	15	152	0,15
4	14	20	19	21	10	16	20	19	13	19	171	0,17
5	18	13	10	13	18	16	22	19	22	18	168	0,17
6	18	27	18	16	14	16	14	14	23	21	181	0,18
Suma	100	100	100	100	100	100	100	100	100	100	1000	1

A partir de los resultados parciales, se calcularán las frecuencias (Figura 3, a)) y, se obtendrá la representación gráfica (Figura 3 b)). El valor promedio es 3.5544 y la varianza

es igual a 3.006. Al calcular el valor χ^2 global se obtiene $\chi^2 = 3,18896 < \chi_{5,0,05}^2 = 11,070$. En este caso, el resultado, a pesar de contar con 1000 simulaciones, es “peor” que el obtenido con sólo 100 simulaciones puesto que el valor obtenido para el estadístico χ^2 es mayor. Este hecho es posible pero, si se repitiera muchas veces la experiencia, no sería lo usual. En general con 1000 simulaciones obtendríamos resultados más ajustados que con sólo 100 simulaciones.

a)

Valor	Frecuencia absoluta	Frecuencia relativa
1	169	0,17
2	159	0,16
3	152	0,15
4	171	0,17
5	168	0,17
6	181	0,18
SUMA	1000	1

b)

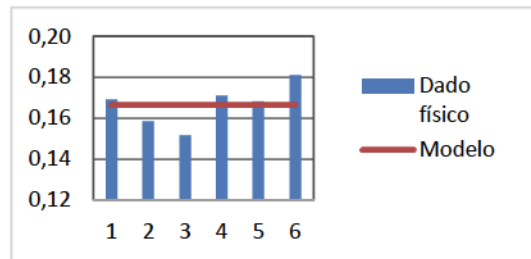


Fig. 3 a) Ejemplo de tabla de frecuencias y b), gráfico de frecuencias relativas obtenidas de la experimentación con el dado físico, con la línea de referencia correspondiente a las probabilidades del modelo teórico.

Etapa 3.-Simulación mediante Excel®

En esta etapa, para generar los valores pseudo-aleatorios uniformes discretos en el intervalo [1,6], utilizaremos una fórmula de Excel que desarrolla el Algoritmo 1 del apartado 1.5. El valor pseudo-aleatorio correspondiente a una variable distribuida uniforme continua en el intervalo (0,1) se obtiene mediante la orden =ALEATORIO(). Si en la casilla A1 introducimos la orden =ALEATORIO(), y en la casilla B1 introducimos la fórmula

=SI(A2<=1/6;1;SI(A2<=2/6;2;SI(A2<=3/6;3;SI(A2<=4/6;4;SI(A2<=5/6;5;6))))

obtendremos en B1, un valor entero en el intervalo [1,6] que, en teoría, sería una realización de una variable aleatoria uniforme discreta en el intervalo [1,6]. Aunque Excel tiene la función =ALEATORIO.ENTRE(1;6) que proporciona directamente un valor entero en el intervalo [1,6], se ha optado por la fórmula expuesta por ser fácilmente generalizable a variables aleatorias discretas no uniformes.

Reproduciendo las fórmulas de las celdas A1 y B1 en las siguientes, hasta llegar a las celdas A1000 y B1000, obtenemos una muestra de 1000 valores enteros correspondientes a una variable aleatoria uniforme discreta en el intervalo [1,6]. En la figura 4 a) y 4 b) se muestran las frecuencias absolutas y relativas y la representación gráfica de las frecuencias relativas obtenidas en la simulación junto a las probabilidades teóricas del modelo, respectivamente.

a)

Valor	Frecuencia absoluta	Frecuencia relativa
1	175	0,175
2	159	0,159
3	156	0,156
4	175	0,175
5	174	0,174
6	161	0,161
SUMA	1000	1

b)

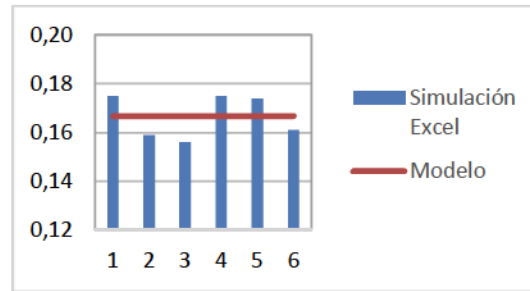


Fig. 4 a) Ejemplo de tabla de frecuencias y b), gráfico de frecuencias relativas obtenidas de la simulación con Excel, con la línea de referencia correspondiente a las probabilidades del modelo teórico.

Para la simulación expuesta, la media es igual a 3,485 y la varianza 2,907. Por otra parte, se tiene el valor obtenido para $\chi^2=2.384 < \chi_{5,0.05}^2=11.070$, lo cual indica un buen ajuste al modelo teórico, a un nivel de confianza del 95%.

Etapa 4.-Simulación mediante MATLAB®

Para generar los 1000 valores pseudo-aleatorios uniformes discretos en el intervalo [1,6] mediante Matlab utilizaremos el script:

```
A=unidrnd(6,1000,1);
tabulate(A)
media=mean(A)
varianza=var(A)
table=tabulate(A);
```

Que proporciona la media aritmética (3.4750) la varianza (2.8903) y la tabla de frecuencias absolutas y relativas, éstas últimas en porcentaje.

```
media =
    3.4750
varianza =
    2.8903
>> Value Count Percent
      1    171    17.10%
      2    164    16.40%
      3    165    16.50%
      4    180    18.00%
      5    159    15.90%
      6    161    16.10%
```

En este caso, el valor obtenido para $\chi^2=1.784 < \chi_{5,0.05}^2=11.070$, también indica un buen ajuste al modelo teórico. El gráfico de frecuencias relativas correspondiente es el expuesto en la figura 5

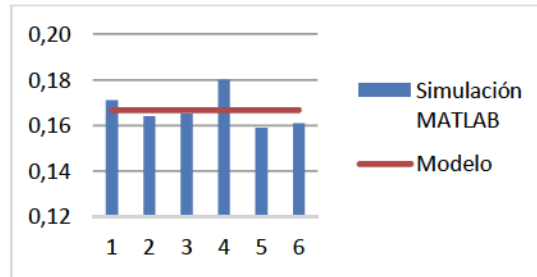


Fig. 5 Gráfico de frecuencias relativas obtenidas de la experimentación con Matlab, con la línea de referencia correspondiente a las probabilidades del modelo teórico.

Al final, todos los resultados se resumen en una tabla conjunta (tabla 3)

Tabla 3: Resumen con los resultados y estadísticos correspondientes a los tres simuladores.

	Modelo	Dado físico	Simulación EXCEL	Simulación Matlab
1	$1/6=0.1\hat{6}$	0,17	0,18	0,17
2	$0.1\hat{6}$	0,16	0,16	0,16
3	$0.1\hat{6}$	0,15	0,16	0,17
4	$0.1\hat{6}$	0,17	0,18	0,18
5	$0.1\hat{6}$	0,17	0,17	0,16
6	$0.1\hat{6}$	0,18	0,16	0,16
Valor esperado (μ)	3.5	3.55	3.49	3.48
Varianza (σ^2)	$2.91\hat{6}$	3.01	2.91	2.89
χ^2	0	3.189	2.384	1.784

La revisión de la tabla 3 indica que, en nuestra experiencia, la simulación que más se acerca al modelo teórico es la que proporciona la simulación Matlab, la cual presenta un valor menor de χ^2 . No obstante, todos los valores de χ^2 obtenidos son mucho menores que $\chi^2_{5,0.05} = 11.070$, por lo que podemos asegurar a un nivel de confianza del 95% que todos los simuladores se ajustan a la distribución de una variable aleatoria uniforme discreta [1,6]. En cierta forma hemos certificado la validez tres métodos para simular el comportamiento de un dado perfecto y, no lo olvidemos, de las crías vivas de una camada (RI).

Etapa 5.-Trabajo propuesto: Certificando la validez de un simulador disponible en Internet.

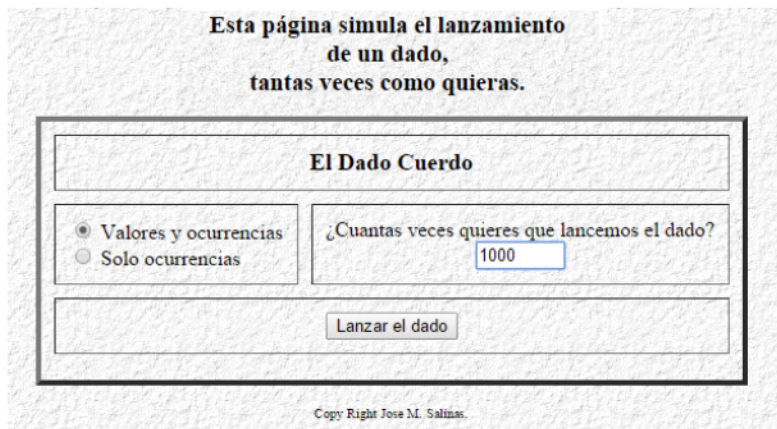
En Internet podemos encontrar múltiples páginas que ofrecen la posibilidad de generar valores pseudo-aleatorios correspondientes a distintas distribuciones (Radikewl, 2016). Una de ellas es la que nos proporciona la página personal de José María Salinas Martínez de Lecea, en la web de la Universidad de Granada, donde entre diversas opciones tenemos la que se titula *Simulador de lanzamientos de un dado correcto*.

El objetivo de esta etapa es certificar la validez del simulador web como generador de valores pseudo-aleatorios correspondientes a una distribución uniforme discreta en el intervalo [1,6].

Trabajar con el simulador es sencillo. Aparece una pantalla como la que se expone en la figura 6 a), donde se seleccionará la opción *Valores y ocurrencias*, y que queremos que el “dado” se lance 1000 veces. Un posible resultado se muestra en la figura 6 b), donde se

observa una vista parcial de las ocurrencias correspondientes a las 1000 simulaciones y el resumen de los resultados.

a)



b)

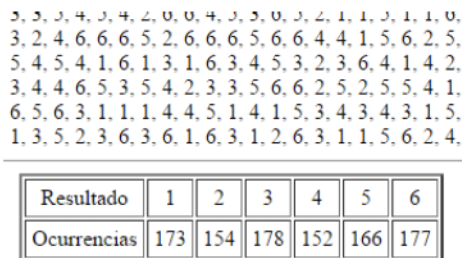


Fig. 6 a) Opciones del generador web. b) Resultados al ejecutar una simulación con 1000 repeticiones.

Se propone a los alumnos que certifiquen, con un nivel de confianza del 95%, si el simulador web ofrece resultados que se ajustan a una distribución uniforme discreta en el intervalo [1,6].

4. Resultados de la experiencia en el aula

La experiencia se ha desarrollado parcialmente en el primer cuatrimestre del curso 2015-2016, en la asignatura Instrumentos de Estadística y Simulación de segundo curso del Grado en Ciencias Ambientales, en la Escola Politècnica Superior de Gandia-Universitat Politècnica de València. Para un número total de alumnos de 52 alumnos, distribuidos en dos grupos, se ha diseñado y puesto en marcha una práctica en aula informática, tomando como base lo explicado en clase magistral de teoría y material en formato pdf disponible en la plataforma PoliformaT/Recursos (etapas 2, 3, 4, y 5). El material complementario necesario se ha proporcionado en la propia aula informática antes de iniciar la práctica: por una parte el dado físico y por otra parte ficheros con instrucciones y rutinas Matlab auxiliares, adjuntados a la actividad propuesta utilizando la herramienta PoliformaT/Tareas. Los alumnos han trabajado en grupos de dos personas y han tenido que cumplimentar una ficha donde, secuencialmente, tenían que trasladar los resultados de la experimentación con el dado físico, con el generador basado en la fórmula de Excel y con el generador de valores pseudo-aleatorios de Matlab. Por último tenían que verificar si los resultados

indicaban que podemos aceptar las simulaciones como adecuadas para representar un dado perfecto. La actividad ha tenido que ser muy guiada en la parte correspondiente a la simulación con Excel y con Matlab, a pesar de que teóricamente los alumnos debían estar familiarizados con estos programas. Por otra parte se han tenido que repasar conceptos ya explicados en clase, que tenían que haber trabajado previamente en casa.

Como primer aspecto a tener en cuenta, cabe destacar el interés mostrado en el desarrollo de la actividad. Lejos de centrarse únicamente en cumplimentar la ficha, en general los alumnos se han implicado activamente en el trabajo, posiblemente influenciados positivamente por la actividad manipulativa de obtener los primeros resultados a partir de un dado físico.

Un compendio de los resultados, transcritos a una ficha en formato de Microsoft Word, así como los ficheros Excel utilizados en el trabajo (que cada grupo debe subir a la correspondiente tarea de PoliformaT) han servido como elementos para evaluar la actividad.

En un 67% de los casos, lo que presentan los alumnos mantiene un orden adecuado, prueba de que el grupo controlaba la situación y entendía aquello sobre lo que se estaba trabajando, con lo que se puede concluir que han adquirido las competencias objetivo. En el resto de los casos, no podemos valorar como satisfactorio el resultado por diversas razones, como el desorden en el documento Word presentado, errores de cálculo con resultados imposibles, trabajo incompleto, ... Estimamos que en el 33% de los casos las competencias-objetivo no se han adquirido o se han adquirido parcialmente.

Una valoración global del desarrollo de la experiencia y de los resultados nos llevan a la conclusión de la necesidad de reforzar conceptos e ideas antes del desarrollo de la actividad práctica presencial. Esta realidad es la que motiva la conveniencia de introducir en el proceso la Etapa 1, para la cual se propone la metodología flip teaching.

En el curso 2016-17 se incorporarán las etapas 1 y 6 a la experiencia. Los resultados del curso 2015-16 indican que, con la incorporación de la etapa 1, se podrá dar feedback a los alumnos para reforzar conceptos e ideas antes de la realización de la práctica presencial. Por otra parte, la etapa 6 permitirá reforzar competencias relacionadas con la capacidad de trabajo autónomo de los alumnos.

5. Conclusiones

La comprensión del concepto de variable aleatoria no es algo inmediato. El hecho de manejar una variable, asociada a un experimento aleatorio, que sabemos qué valores puede tomar, pero no se pueden predecir aunque siguen cierta ley, va asociado a crear un constructo cognitivo que permita establecer una correspondencia entre la frecuencia relativa para los diversos resultados posibles al repetir muchas veces un experimento aleatorio con la probabilidad teórica para cada uno de dichos resultados.

La experiencia propuesta facilita al alumnado aprender a establecer la relación entre experimento aleatorio y modelo teórico estadístico. Por otra parte, también permite que los alumnos puedan asimilar e interiorizar la idea de cómo valorar el nivel de ajuste de una

muestra, producto de una simulación, a una distribución teórica, lo cual tampoco es inmediato. La tendencia natural de los alumnos es reducir la formación e información proporcionada en la asignatura a una serie de reglas prácticas en base a las que actuar y resolver problemas sin profundizar en el significado de lo que se está haciendo. Con el desarrollo de esta experiencia se consigue romper en parte esta tendencia, gracias al trabajo con el dado físico y el software, y al uso del test Chi-cuadrado para los distintos simuladores, identificando la superación del test con el hecho de disponer de una herramienta que permite certificar la validez de un dado.

6. Referencias

BLANCO BLANCO, A. (2008). Una revisión crítica de la investigación sobre las actitudes de los estudiantes universitarios hacia la estadística. *Revista Complutense de Educación*, 19(2), 311–330.

COLABORADORES DE WIKIPEDIA, *Generador de números aleatorios*, Wikipedia, La enciclopedia libre, <https://es.wikipedia.org/w/index.php?title=Generador_de_n%C3%BAmoros_aleatorios&oldid=88483665> [Consulta 3 de febrero de 2016]

COLABORADORES DE WIKIPEDIA, "*Generador de números pseudoaleatorios*", Wikipedia, La enciclopedia libre, <https://es.wikipedia.org/w/index.php?title=Generador_de_n%C3%BAmoros_pseudoaleatorios&oldid=88579925> [Consulta 3 de febrero de 2016].

ESCALANTE GÓMEZ, E. (2008). Actitudes de alumnos de posgrado hacia la estadística aplicada la investigación. *Encuentro 2010/ Año XLII*, nº 85, 27-38.

"Estadística útil. ¿Qué es una prueba de bondad de ajuste?". *Youtube* <<https://www.youtube.com/watch?v=nxXVwOp4444>> [Consulta: 1 de noviembre de 2016]

"FisicayMates. Variables aleatorias discretas y continuas | Estadística UNED". *Youtube* <https://www.youtube.com/watch?v=n0T_HcJ7oak> [Consulta: 1 de noviembre de 2016]

"FisicayMates. Variable Aleatoria Discreta | Funcion de probabilidad y de distribucion". *Youtube* <<https://www.youtube.com/watch?v=naEqsDvkIXs>> [Consulta: 1 de noviembre de 2016]

INSTITUTO TECNOLÓGICO DE CHIHUAHUA. *Prueba chi-cuadrada para la bondad del ajuste*. <<http://www.itch.edu.mx/academic/industrial/estadistica1/cap04b.html>> [Consulta: 1 de noviembre de 2015] [Leticia de la Torre]

KAY, S.M. (2006). *Intuitive Probability and Random Processes using MATLAB*. New York; Springer

RADIKEWL *Generadores Online Gratis números aleatorios y "ganadores" Estrategias de la Lotería*, < <http://www.radikewl.com/2252801.html>> [Consulta: 1 de noviembre de 2015] [sin autoría reconocida]

UNIVERSIDAD DE GRANADA. *Herramientas Estadísticas*, <<http://www.ugr.es/~jsalinas/herramar.htm>> [Consulta: 1 de noviembre de 2015] [José María Salinas Martínez de Lecea]